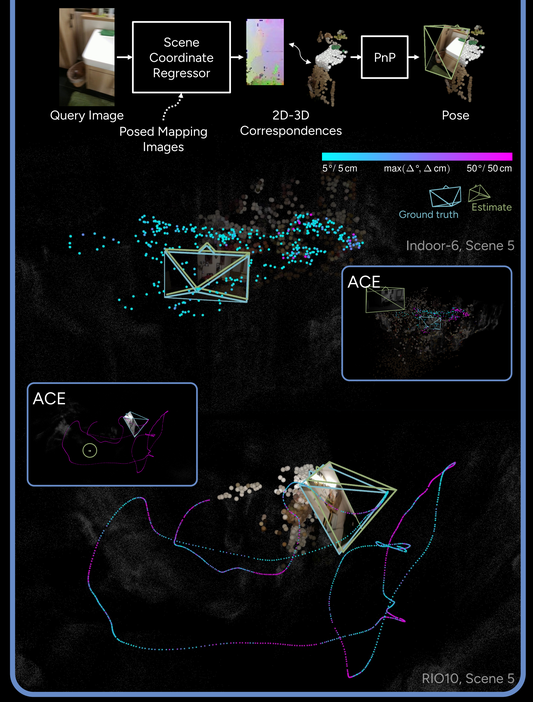
ACE-G Improving Generalization of Scene Coordinate Regression Through Query Pre-Training

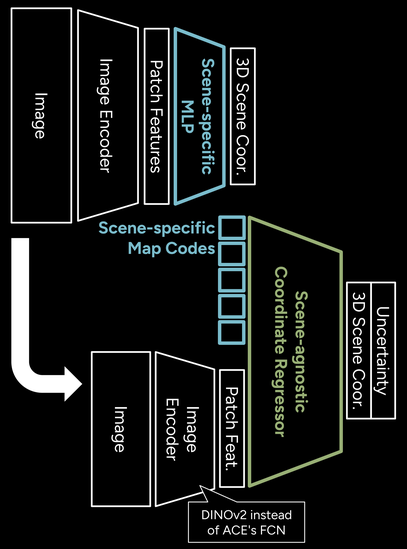
Leonard Bruns Axel Barroso-Laguna Tommaso Cavallari Áron Monszpart Sowmya Munukutla Victor Adrian Prisacariu Eric Brachmann

NIANTIC SPATIAL

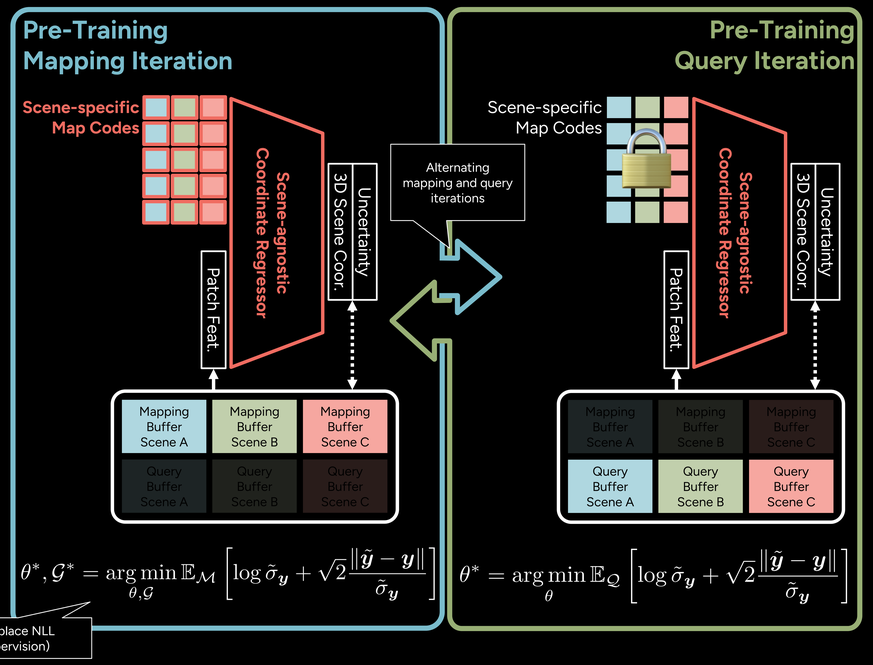
Visual Relocalization



Split into **scene-specific map codes** and **scene-agnostic coordinate regressor** ...



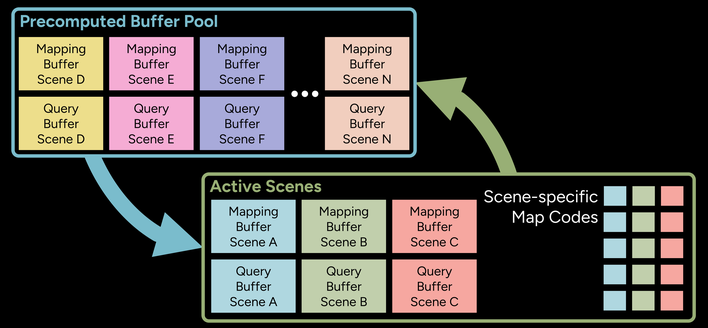
... and **pre-train** explicitly to generalize from **mapping** to **query** conditions ...



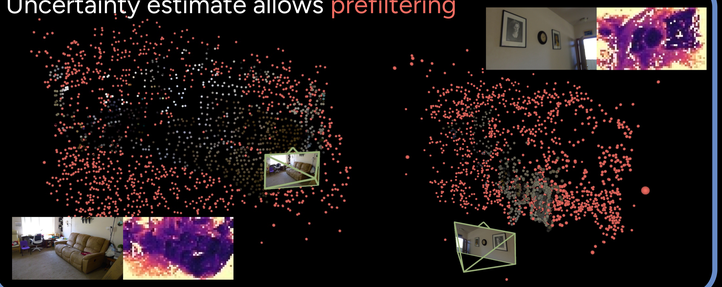
$$\theta^*, \mathcal{G}^* = \arg\min_{\theta,\mathcal{G}} \mathbb{E}_{\mathcal{M}} \left[\log \tilde{\sigma}_{\boldsymbol{y}} + \sqrt{2} \frac{\|\tilde{\boldsymbol{y}} - \boldsymbol{y}\|}{\tilde{\sigma}_{\boldsymbol{y}}} \right]$$

$$\theta^* = \arg\min_{\theta} \mathbb{E}_{\mathcal{Q}} \left[\log \tilde{\sigma}_{\boldsymbol{y}} + \sqrt{2} \frac{\|\tilde{\boldsymbol{y}} - \boldsymbol{y}\|}{\tilde{\sigma}_{\boldsymbol{y}}} \right]$$

... using a continuously changing set of scenes.



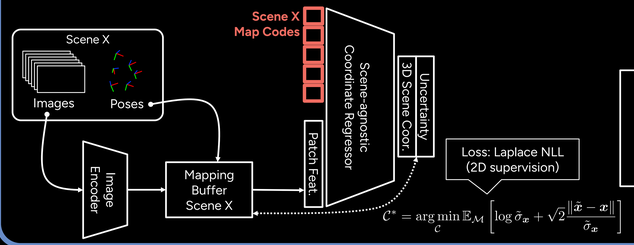
Uncertainty estimate allows prefiltering



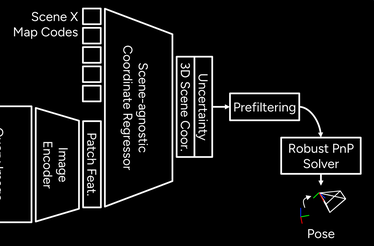
Why? Scene-specific scene coordinate regressors do not generalize well to changes in ...

Lighting Viewpoint Objects

Learn a map code (~10MB) for a new scene...



$$\mathcal{C}^* = \arg\min_{\mathcal{C}} \mathbb{E}_{\mathcal{M}} \left[\log \tilde{\sigma}_{\boldsymbol{x}} + \sqrt{2} \frac{\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|}{\tilde{\sigma}_{\boldsymbol{x}}} \right]$$



... and localize an image within.

Quantitative Results

	(cm / °)	Map. Time	Map Size	Indoor-6 Scene 5	Indoor-6 Avg.	RIO10 Scene 5	RIO10 Avg.
Non SCR	Reloc3r	< 5 min.	~500 MB	22 / 1.4	9 / 0.8	35 / 3.8	47 / 9.4
	MASt3R+Kapture	5-10 h	~5 GB	4.3 / 0.7	2.5 / 0.5	68 / 22.9	N/A
SCR 25 min.	GLACE	25 min.	6 MB	6.1 / 0.9	**3.5 / 0.6**	127.8 / 45.6	114 / 33.4
	Ours (25 min.)	25 min.	12 MB	**5.1 / 0.8**	3.7 / 0.7	**17.6 / 6.4**	**37.0 / 12.1**
SCR 5 min.	ACE	5 min.	4 MB	21.6 / 4.0	11.0 / 1.8	915.9 / 95.0	358.4 / 58.7
	ACE w/ DINOv2	5 min.	4 MB	7.5 / 1.2	5.6 / 1.0	118.2 / 10.6	83.8 / 15.9
	Ours (5 min.)	5 min.	12 MB	**5.9 / 1.0**	**4.5 / 0.8**	**20.3 / 7.4**	**41.1 / 13.8**